DOCUMENT RESUME

ED 219 452 TM 820 480

AUTHOR Diamond, Esther E.
TITLE Increasing Public Understanding of School Testing Programs.
PUB DATE Mar 82
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Communication (Thought Transfer); Minimum Competency Testing; *Public Opinion; Student Placement; *Testing; Testing Problems; *Testing Programs; Test Interpretation
IDENTIFIERS *Test Publishers

ABSTRACT

Increasing public understanding of school testing programs is seen as the joint responsibility of test publishers and schools, and also as a primary responsibility on the part of test publishers to increase the schools' understanding of tests. The most common areas of misunderstanding about tests are addressed: purpose of the testing program; the myth of test score infallibility; meanings of different types of test scores; defining minimum competency and setting standards; using tests to classify and place exceptional children; bias in testing; and tests as only one source of information. The importance of continuous communication, from the time the decision to test is first made, is emphasized, and various aids to communication are suggested. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219452

Increasing Public Understanding of School Testing Programs[1]

Esther E. Diamond
Science Research Associates

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E.E. Diamond

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: Public Understanding of Tests

[1]Paper presented at the annual meeting of the American Educational Research Association, New York City, March 1982, as part of the symposium Problems, Needs, New Directions in Testing: A Test Directors/Test Publishers Dialogue. The views expressed here are the author's and do not necessarily represent those of Science Research Associates.

This paper may not be distributed or reproduced in any way without prior written permission of the author.

TM 820 480

Abstract

Increasing public understanding of school testing programs is seen as the joint responsibility of test publishers and schools, and also as a primary responsibility on the part of test publishers to increase the schools' understanding of tests. The most common areas of misunderstanding about tests are addressed: purpose of the testing program; the myth of test score infallibility; meanings of different types of test scores; defining minimum competency and setting standards; using tests to classify and place exceptional children; bias in testing; and tests as only one source of information. The importance of continuous communication, from the time the decision to test is first made, is emphasized, and various aids to communication are suggested.

Increasing Public Understanding of School Testing Programs

Esther E. Diamond

Science Research Associates

Increasing public understanding of school testing programs is the joint responsibility of test publishers and schools, but it is also a primary responsibility of publishers to increase the schools' understanding of the publishers' tests--their purpose, content, uses and possible abuses, advantages and limitations, and psychometric characteristics. Only when this has been accomplished can the schools inform the public and solicit their confidence and support.

What are some of the most common areas of misunderstanding about tests that need to be addressed?

1. Purpose of the Testing Program

The public is often unclear about the specific purpose of the school's testing program. They need to know that the overriding purpose is to improve student learning through instruction. Standardized achievement tests and mastery and criterion-referenced tests tailor-made to the school curriculum can pinpoint individual student and group weaknesses in specific skills and content areas. Diagnostic tests can shed light on specific learning difficulties. All other uses (for example, evaluation of the progress of an individual student, a class, a grade, or a school system; assignment of students to appropriate instructional programs; and evaluation of an experimental intervention)

are related but secondary to the improvement of learning. (Not included in this discussion are tests such as interest inventories and other assessments designed to increase student self-understanding, help students make personal decisions, or help the school understand student feelings and behavior.)

Schools and the public at large need to understand the differences between various types of tests as well as their similarities, and the reason for including them in the school testing program. It should be made clear, too, that tests cannot be used as a measure of characteristics or qualities that they are not designed to measure. For example, one cannot use the 99th percentile on an achievement test, or a grade equivalent score two years above grade level to identify geniuses or students who possess that elusive quality known as creativity.

2. The Myth of Test Score Infallibility

Too often, a test score on a standardized test is considered a very precise point on a given scale. The public seems to expect test scores to be 100 percent free from error--whether in the sampling of the population to be tested or of the curriculum, or in the student's own testing behavior at any given point in time. Given the less-than-laboratory situation in which achievement tests, for example, must be developed and standardized, this is a highly unrealistic expectation. Yet the concept of measurement error is one of the most difficult for the public to accept. It may help to provide information about how test content was determined, how representative it is of national or

local curriculums, how the standardization population was selected, and how the raw scores were transformed into other types of scores. Reported scores for most tests indicate the band within which "true" scores are expected to lie, and this concept needs to be communicated as graphically as possible. A common device is using an analogy from sports--how a player's performance can vary from event to event, and how diurnal variations such as personal well-being, weather, and other such factors can affect his or her playing and, consequently, average performance.

3. Meanings of Different Types of Test Scores

How many of us have had the experience of receiving an inquiry from a parent whose third-grade child has a grade equivalent (G.E.) score of 5.2 and who wants the school to double-promote the child because he or she "is reading at the fifth-grade level"? More serious is the case of a school district that uses grade equivalents to establish minimum competence and marks for retention in a grade students who have not achieved a score at the median for that grade. They do not understand that 50 percent of the children in the grade, by definition, are below the median--the point represented by the G.E. Both schools and the public also need to understand that the grade-equivalent scale is not an equal-interval scale, and that sometimes very little actual change in performance can result in a grade equivalent difference of a year or more.

Percentiles can also be troublemakers if it is not understood that a percentile is a relative and not an absolute score, and that percentile standings must be considered in the light of the reference population.

Sometimes, too, a school district decides to establish a percentile as a minimum cutoff score--overlooking the fact that in any population, even in a population of geniuses with small differences of ability among them, 25 percent would be below the 25th percentile, 35 percent below the 35th percentile, and so on.

Finally, a potential troublemaker is a score that has just recently made its appearance in score reports for several widely used tests--the predicted, anticipated, or expected score based on a test of school ability and possibly other factors such as age, sex, and SES. Where such scores are provided, they should be interpreted with utmost caution. The recently issued report by the Committee on Ability Testing (1982) maintains that all of the tests used in schools, including aptitude and general ability tests, are measures of currently developed competence and should be considered as providing estimates of students' present level of functioning in some domain.

4. Defining Minimum Competency and Setting Standards

The subject of cutoff scores for determining minimum competency has been touched on briefly earlier in this paper, but the point I'd like to emphasize here--one that has caused considerable public alarm in a number of states and districts--is the practice of establishing cutoff scores without empirical evidence that students just below the cutoff cannot handle work at the next grade level or that those just above the cutoff can handle it. In some cases, too, students retained in grade have not been given remedial work, so that their chances of ever performing at grade level were minimal. Sometimes held-back

students are relegated to custodial programs, where they stagnate in an academic never-never land.

The public should be informed of the basis for cutoff scores, and it should be informed of plans to bring up to grade level the performance of students who are held back. Retention without appropriate remediation overwhelmingly increases the odds against the student and must surely baffle if not embitter the public.

In addition, as the Committee on Ability Testing (1982) points out, schools should make a continuing effort to check on the individual soundness of any plans for grouping or classifying students. Reviewing minimum competency plans and decisions would seem to apply here.

5. Using Tests to Classify and Place Exceptional Children

Court cases resulting from disproportionate assignment of racial minorities to special education classes have created a strong antitesting climate on the part of various segments of the public (Bersoff, 1981). As Reschly (1981) points out, however, standardized tests are only a small part of the overall problem of disproportionate classification and placement procedures for exceptional children. While reform in assessment, classification, and placement procedures is needed, including the use of multiple measures, the real problem is ensuring that classification leads to necessary and effective services. Overall, Reschly argues, available data suggest that IQ tests (tests of academic ability) either have a neutral effect on disproportionality or actually reduce the degree of overrepresentation that would result from teacher referral alone.

6. Bias in Testing

Public concern about bias in testing is particularly volatile in communities characterized by large ethnic minorities who score in the lower part of the score distribution on standardized achievement tests. The public needs to know what the tests are testing, what is considered bias in testing, and what has been done to minimize bias in the tests being used. Equally important, the public needs to know that much of the bias attributed to tests is actually the bias that exists in society and possibly in their own communities. To a limited extent, this knowledge could encourage cooperation between the public and the schools to achieve greater equality of educational opportunity at the local level.

7. Tests as Only One Source of Information

Finally, the public needs to be reassured that a single test score will not be used to make important decisions about students--graduation from school, promotion to the next grade, assignment to various instructional programs, and admission to college. This assurance must, of course, be given in good faith; schools must meticulously police and review the ways in which they use tests, and the extent to which test results are looked at in the light of all other relevant information about a student or group of students. Such information should include teacher observation of classroom performance and overall behavior; out-of-school learning opportunities and experiences; attendance record; emotional and physical factors that might affect performance, including test performance; self-concept; motivation; socioeconomic status; and

dominant language in the home.

Increasing Public Understanding a Continuous Process

Increasing public understanding of the school testing program should be a continuous process. It must begin, not when test scores have been received, but when the decision to test is first made. The public should know what tests will be given, and when and why. They should know something about the tests themselves and how the results will be used. Handouts describing the tests and containing sample items are sometimes available from publishers. When test results are available, school personnel who will be involved in communicating them both within the school and to the public must understand them and be able to present and interpret them clearly.

Parents can be a particularly important ally and source of support in the attainment of mutually agreed-on goals for their children. They need to know the implications of the test results for their child's progress and they need reassurance that the results will be used constructively. Elaborate audio-visual devices may be helpful, but they are not absolutely necessary. There should, however, be such devices as flipcharts or handouts illustrating the major types of scores and what they mean; a question-and-answer sheet concerning many of the popular misconceptions about tests and test scores; information about appropriate and inappropriate comparisons with other schools, systems, states, and regions, as well as with the country as a whole; and a glossary containing terms that will be used

in explaining the testing program--e.g., norm group, percentile, raw score, standard score, score range. Again, publishers often provide ancillary materials that can help the school in this effort.

Providing the press with information about the testing program from its beginning until the end is, as every school district knows, particularly important. Lack of appropriate information and knowledge can cause stories appearing in the press to undo much of the work that has been done to increase public understanding of the school's testing program.

References

Bersoff, D.N. Testing and the law. American Psychologist, 1981, 36, 1047-1056.

Committee on Ability Testing. Ability testing: Uses, consequences, and controversies, Part I. Washington, D.C.: National Academy Press, 1982.

Reschly, D.J. Psychological testing in educational classification and placement. American Psychologist, 1981, 36, 1094-1102.